**Another attempt (Before exam):**

1. pnorm(6, mean = 10, sd = 4)
2. mean = 10 + 5 \* -1

pnorm(3, mean = mean, sd = 4, lower.tail = FALSE)

1. mean = 10 + 5 \* -2

sd = 4

pnorm(3, mean = mean, sd = sd, lower.tail = FALSE)

1. model = lm(eruptions ~ waiting, data = faithful)

summary(model)$coefficients["(Intercept)", "Estimate"]

1. model = lm(eruptions ~ waiting, data = faithful)

summary(model)$coefficients["waiting", "Estimate"]

1. model = lm(eruptions ~ waiting, data = faithful)

predict(model, newdata = data.frame(waiting=80))

1. model = lm(eruptions ~ waiting, data = faithful)

predict(model, newdata = data.frame(waiting=120))

1. range(faithful$waiting)
2. model = lm(eruptions ~ waiting, data = faithful)

sum((faithful$eruptions - predict(model, newdata = faithful))^2)

1. # R square tells us proportion of explained variation by model.

# R squared = 1 - (SSE/ SST)

model = lm(eruptions ~ waiting, data = faithful)

summary(model)$r.squared

1. # Model directly provides this values as Residual standard error.

model = lm(eruptions ~ waiting, data = faithful)

summary(model)$sigma

# Otherwise we could calculate this as sqrt(sum(resisual^2)/n-2)

# sqrt(sum((faithful$eruptions - predict(model, newdata = faithful))^2)/(length(faithful$eruptions) - 2))

1. The same
2. Answers

* A good estimate for the mean of Y when x = 0 is -1.5
* The difference between the y values of observations at x = 10 and x = 9 is 2.3
  + I think this is not correct. Because the fitted model represents mean value of y not actual y. Therefore, we should say that “the difference between mean y values” instead of “difference between the y values”.
* There are observations in the dataset used to fit this regression with negative y values.
  + We cannot say that.

1. Answers

* The SLR model assumes that response variable follows a normal distribution.
  + Yes.
* The SLR model assumes that errors are independent.
  + Yes.
* The SLR model assumes that the relationship between the response and the predictor is linear.
  + Yes.
* The SLR model allows for larger variances for larger values if the predictor variable.
  + I think variance remains same.

1. Suppose you fit a simple linear regression model and obtain beta\_hat\_1 = 0. Does this mean that there is **no relationship** between the response and the predictor?

* We can say there is no significant linear relationship between predictor and response. But we cannot say there is no relationship at all.

First attempt:

# when x = 0 => Y = 10 + epsilon

# This implies that Y will be less than 6 only when epsilon < -4

pnorm(-4, mean = 0, sd = 4)

# when x = -1 => Y = 5 + epsilon

# This implies that Y will be less than 6 only when epsilon > -2

1 - pnorm(-2, mean = 0, sd = 4)

# when x = -2 => Y = epsilon

# This implies that Y will be greater than 3 only when epsilon > 3

1 - pnorm(3, mean = 0, sd = 4)

model = lm(eruptions ~ waiting, data = faithful)

summary(model)

model = lm(eruptions ~ waiting, data = faithful)

summary(model)

6.

intercept = -1.874016

slope = 0.075628

waiting = 80

duration = intercept + slope \* waiting

duration

7.

intercept = -1.874016

slope = 0.075628

waiting = 120

duration = intercept + slope \* waiting

duration

8.

9.

model = lm(eruptions ~ waiting, data = faithful)

sum(model$residuals^2)

10.

model = lm(eruptions ~ waiting, data = faithful)

summary(model)$r.square

11.

model = lm(eruptions ~ waiting, data = faithful)

sd(model$residuals)